Improving the identification rate of data independent label-free quantitative analysis: A proteomics case study on apple fruit

Kim Buts[1], Servaas Michielssens[2], Eisuke Hayakawa[3], Maarten Hertog[1], Bart Nicolai[1], Sebastien Carpentier[3,4]

1. BIOSYST-MeBioS, KU Leuven, Belgium
2. Quantum Chemistry and Physical Chemistry Section, KU Leuven, Belgium
3. SYBIOMA, KU Leuven, Belgium
4. BIOSYST-Crop biotechnics, KU Leuven, Belgium

Complex peptide extracts are troublesome for proper identification and quantification. Fragmentation of the parent ions via data dependent analysis (DDA) is crucial for confident identification. However, DDA results in poor quantification and undersampling. A good alternative is data independent analysis (DIA). DIA assures that all parent ions are thoroughly quantified and fragmented by fast switching between low and high collision energy. With DIA the bottle neck is the identification of peptides. This problem is particularly pronounced in non-model crops such as apple. During long term storage of apple, internal browning can occur inducing unwanted losses. To correlate the proteome to tissue browning, a peptide-based approach is presented, by further adapting our workflow for DIA.

To increase the identification rate of a label free DIA experiment, a DDA based database was constructed and linked to the DIA experiment. In a first phase, DIA and DDA entries were aligned based on their mass to charge ratio (m/z) and retention time (RT); in a second phase, masses of fragmentation ions were compared for each of the linked entries. For the storage experiment Braeburn apples were sampled at harvest and after two weeks, two months and four months of controlled atmosphere storage under brown inducing conditions (2.5% O_2, 3.7% CO_2, 4°C). The DIA experiment comprised 32 samples, for each storage time inside and outside tissue of 4 apples was sampled separately.

After quantitative analysis in Progenesis, 99032 entries were subjected to a multivariate statistical analysis. A partial least square analysis was performed using storage time, brown index and tissue position as y-variables. VIP scores were calculated resulting in a selection of the 5000 most interesting features. Of these features, only 422 peptides (or 8.5%) were identified via DIA.

To increase the identification level a DDA database was built based on 9 subsequent DDA runs of a mixture of the samples used in the DIA. This included2 full scans, followed by 2 exclude runs and 5 include runs. The DDA database contained 17674 peptide sequences, their mass to charge ratios, retention times, charges and masses of the corresponding fragment ions. A DDA and an DIA entry were considered as linked correctly when they were linked in the first phase within the preset m/z and RT window, and when at least 3 fragment masses were linked in the second phase of the process. This resulted in a false discovery rate of 1%. Following this workflow, 1366 extra peptides of the DIA experiment were identified, together with the previous identifications leading to an overall

identification rate of 36%. Further results will be presented focusing on the observed differences in the apple proteome as a function of storage time, brown index and the position of the tissue samples.

To the best of our knowledge, this is the first DIA study on a non-model crop linked to DDA data.

1. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., Geromanos, S. J., 2006. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition Molecular & Cellular Proteomics 5, 144-156.
2. Vertommen, A., Møller, A. L. B., Cordewener, J. H. G., Swennen, R., Panis, B., Finnie, C., America, A. H. P., Carpentier, S. C., 2011. A workflow for peptide-based proteomics in a poorly sequenced plant: A case study on the plasma membrane proteome of banana. J. Proteomics 74, 1218-1229.